What's important to the heart diseases diagnosis platform using deep learning?

- How Openstack & Cyborg could save lives?

Jinghua Gao, Zhenghao Wang (Lenovo Research)
Li Liu (Huawei Canada Research Centre)
2019-04-30

Lenovo 联想

Lenovo

Cyborg's Mission in Democratizing AI

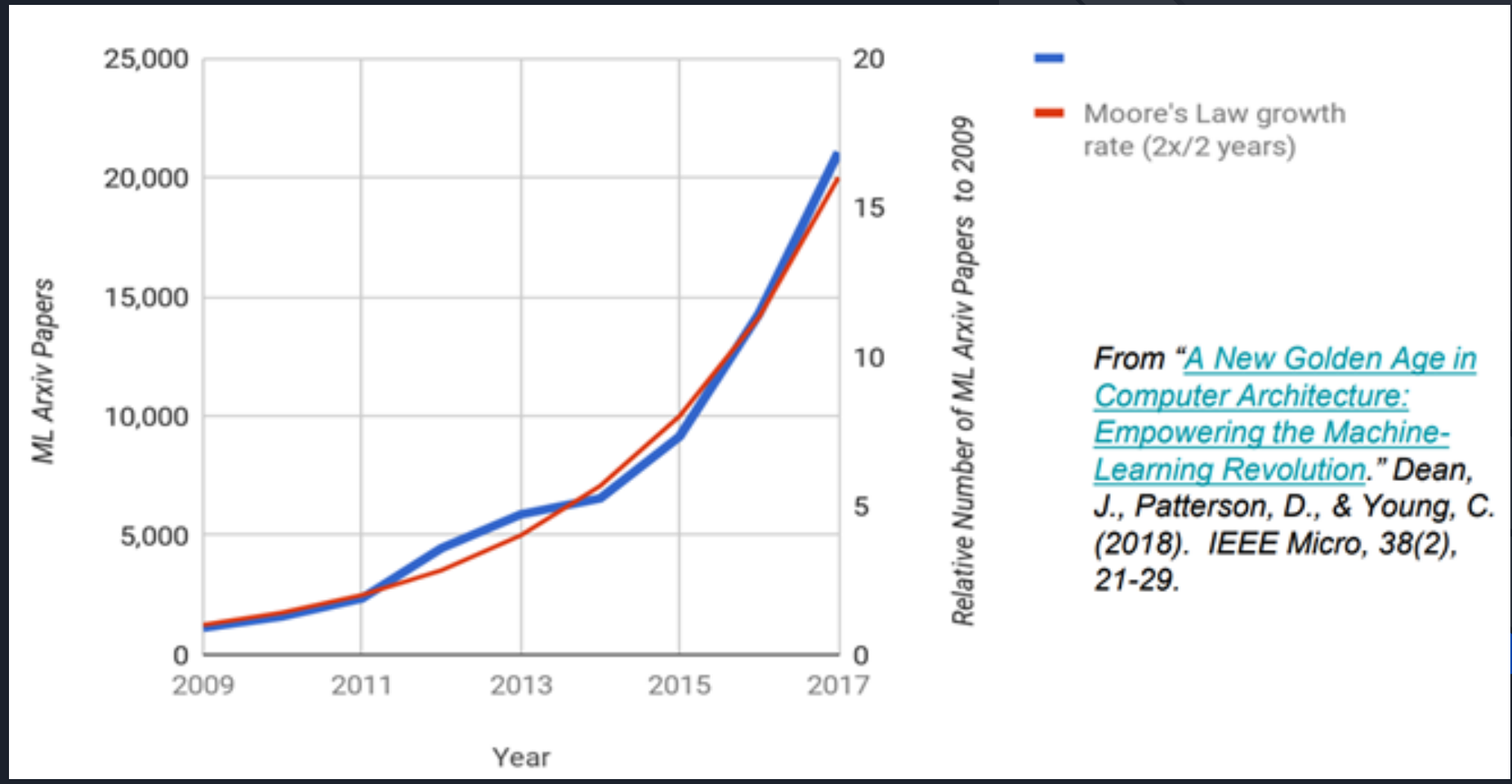
“Ask not what AI can do for you - ask what you can do for AI“

---- Li Liu, Huawei

Outline

- ➔ **Challenges For AI**
- ➔ **Identify The Gap**
- ➔ **Introduce Cyborg Project**
- ➔ **Model Meets Resource**

AI is HOT


25,000
20,000
15,000
10,000
5,000
0
2009
2011
2013
2015
2017
Year
ML Arxiv Papers
20
15
10
5
0
Relative Number of ML Arxiv Papers to 2009
Moore's Law growth rate (2x/2 years)
From "A New Golden Age in Computer Architecture: Empowering the Machine-Learning Revolution." Dean, J., Patterson, D., & Young, C. (2018). IEEE Micro, 38(2), 21-29.

Everyone is talking about democratizing AI

But it can’t be truly done without an open cloud infrastructure

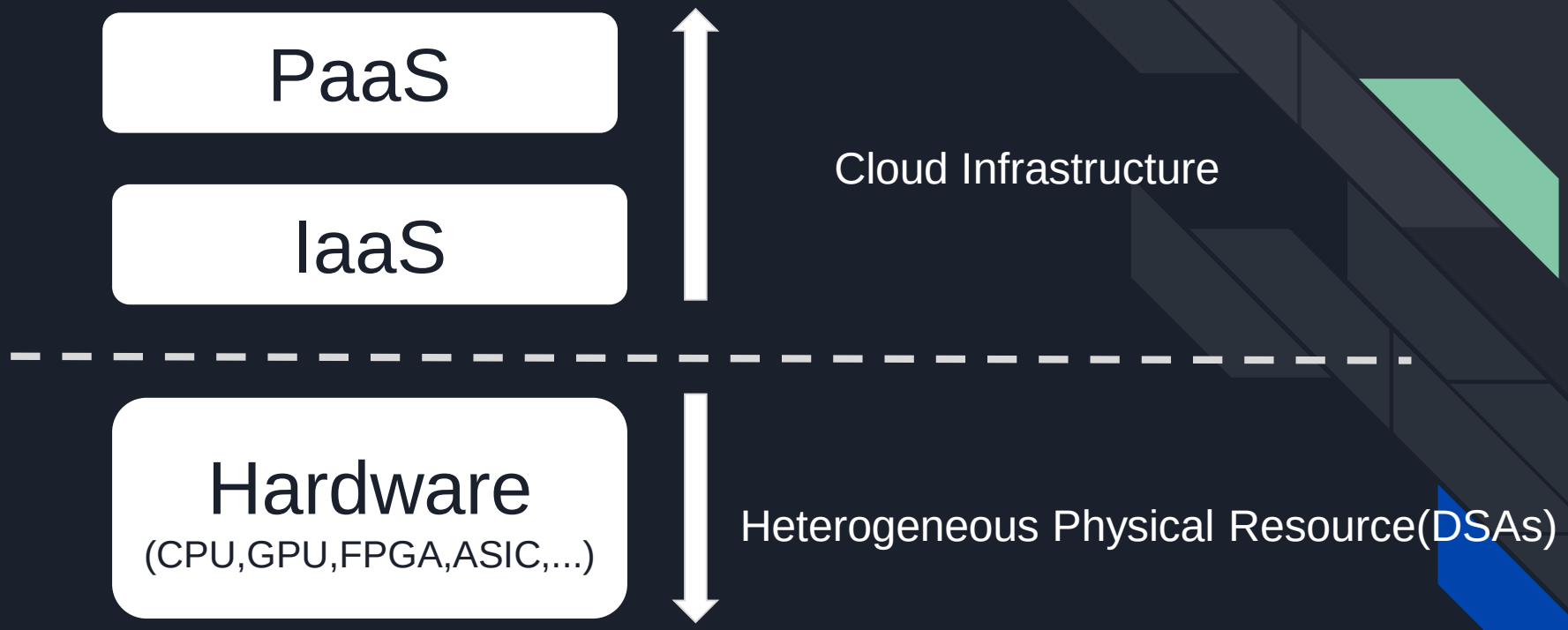
Application

Infrastructure

- Tensorflow, CNTK, Pytorch, Caffe, MXNET, … Basically everything you can find now about major AI related open source projects
- Same goes to majority of the research papers

???

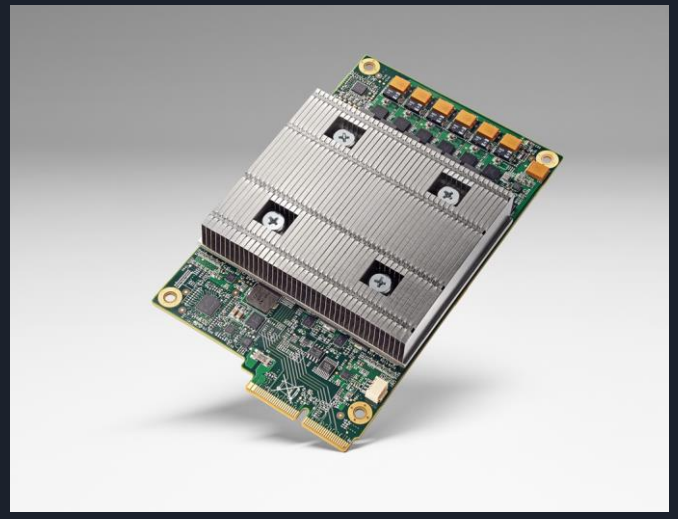
Define a Cloud Infrastructure For AI



Domain Specific Architectures (DSAs)

NPU

Neural network processors for machine learning



GPU

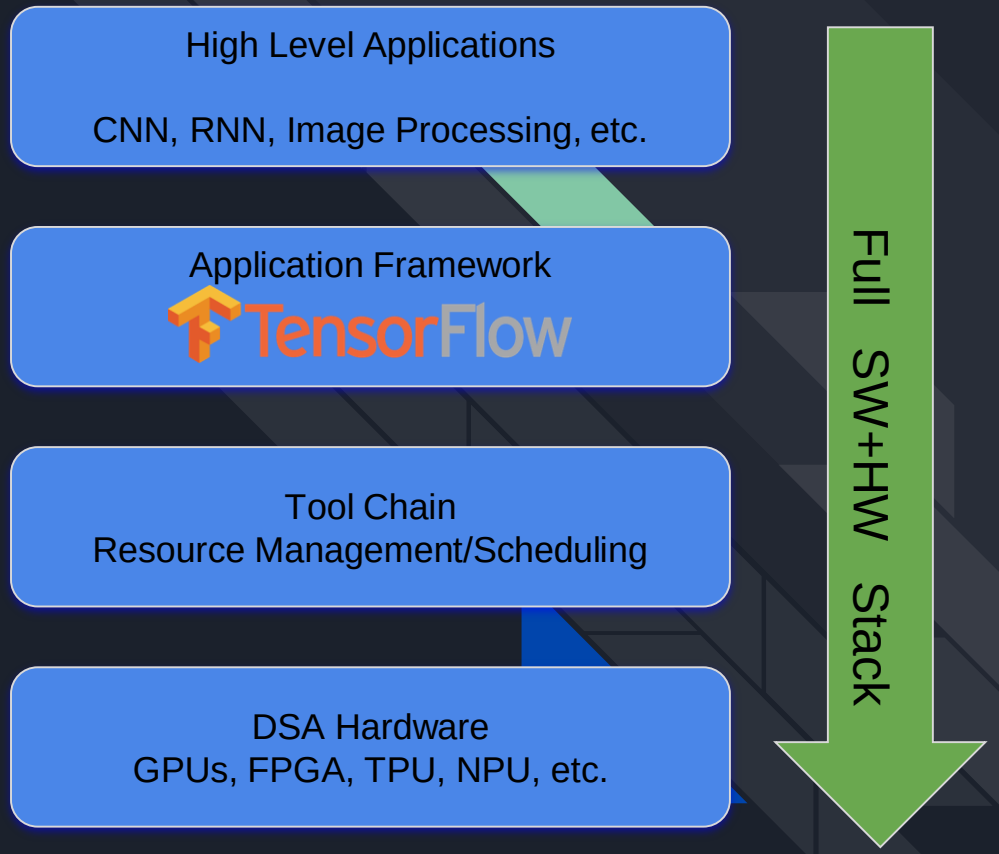
GPUs for graphics, virtual reality, ML

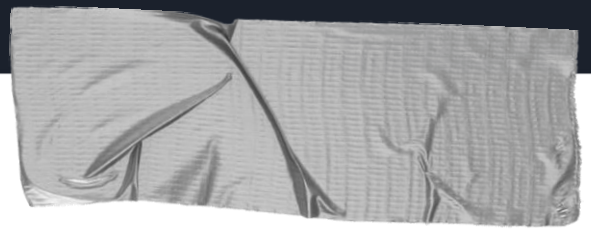
SmartNIC/FPGA

Programmable network switches and hardware

But, what is still missing in the picture?

- DSA is facing fragmentation problems
- Domain Specific Language
  - TensorFlow/Caffe
- Need ecosystem to bridge Cloud, DSA, and DSL

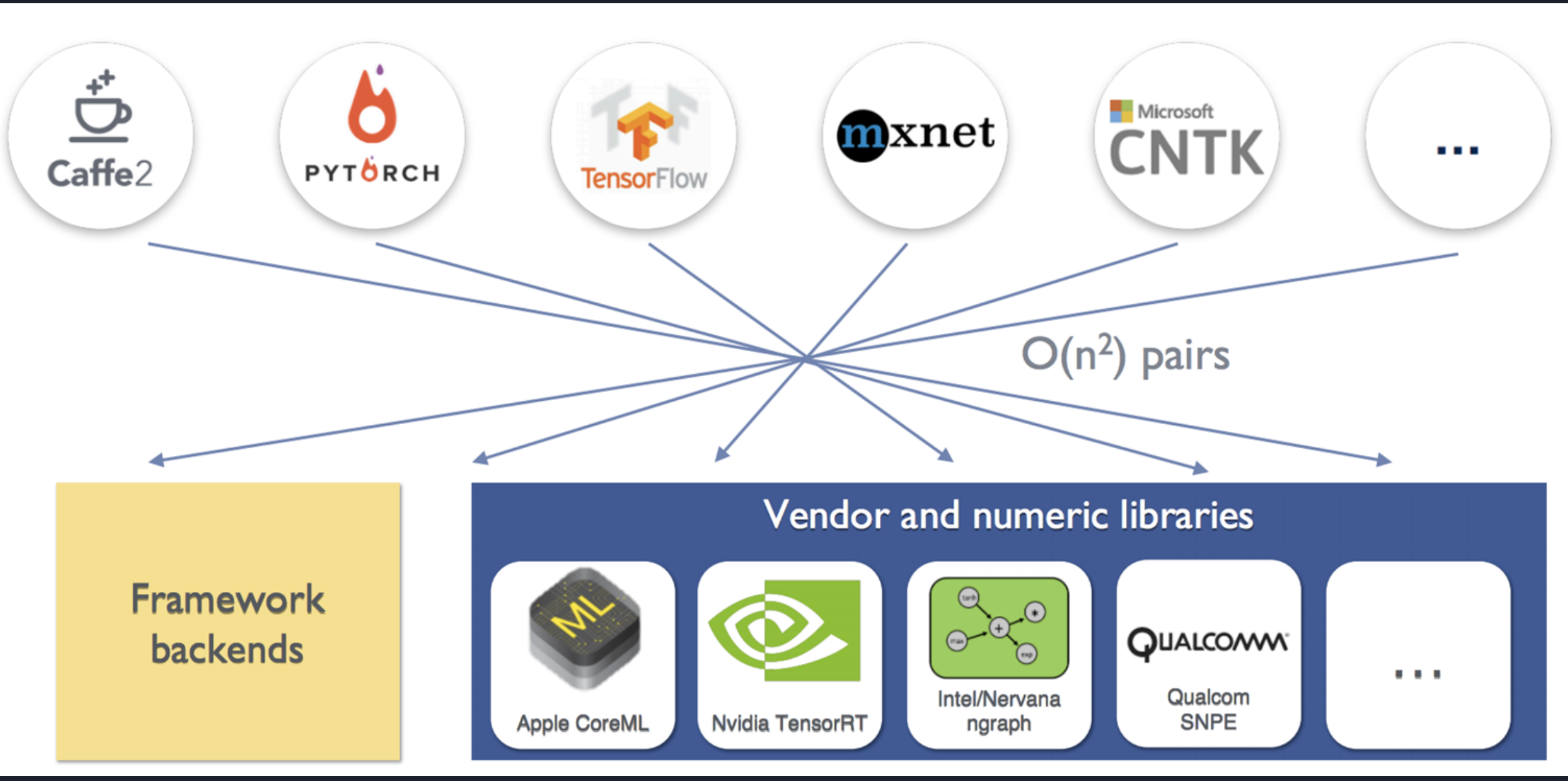




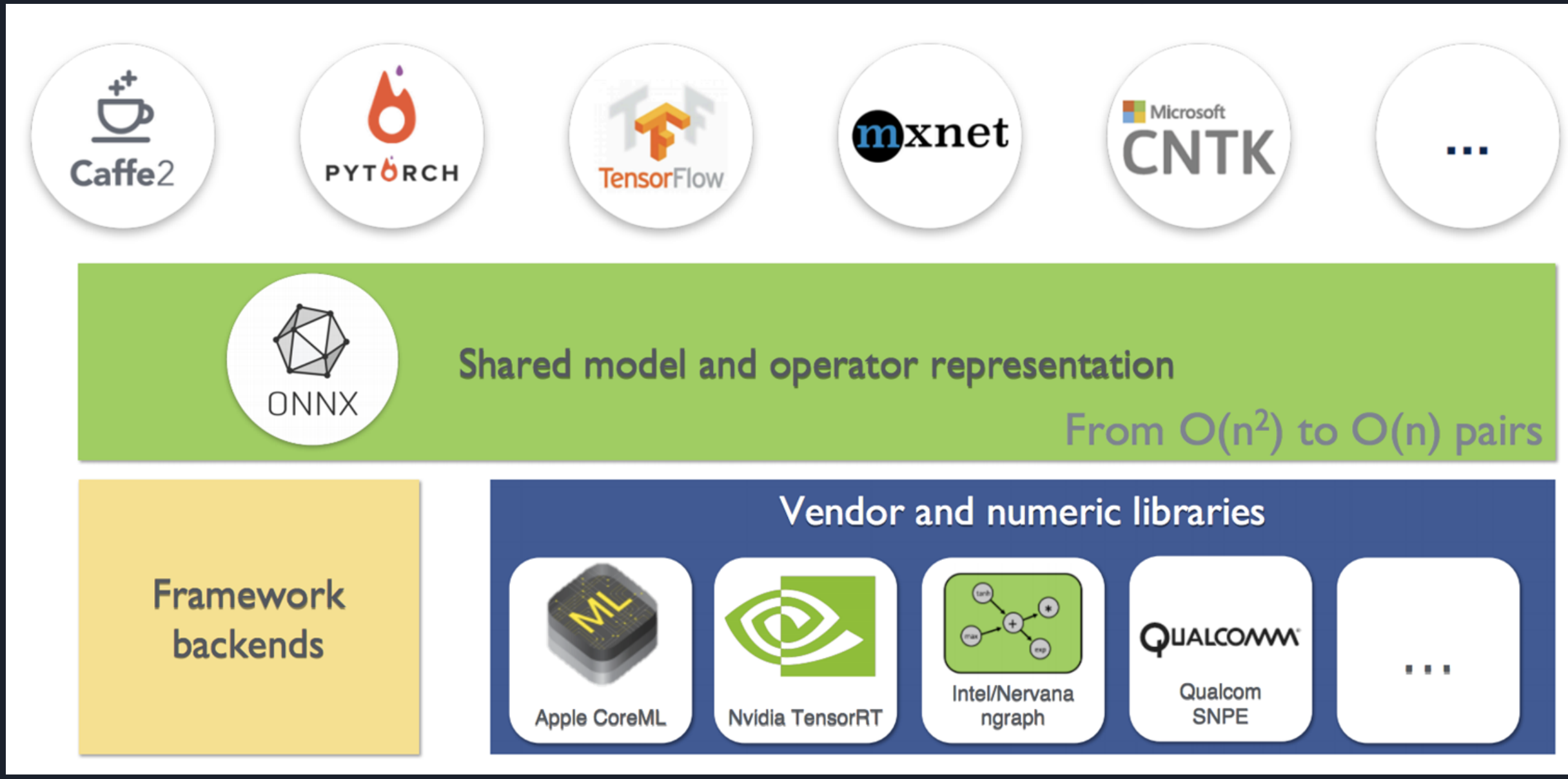
Outline

- ➔ Challenges For AI
- ➔ Identify The Gap
- ➔ Introduce Cyborg Project
- ➔ Model Meets Resource

Deep Learning Models Zoo ...

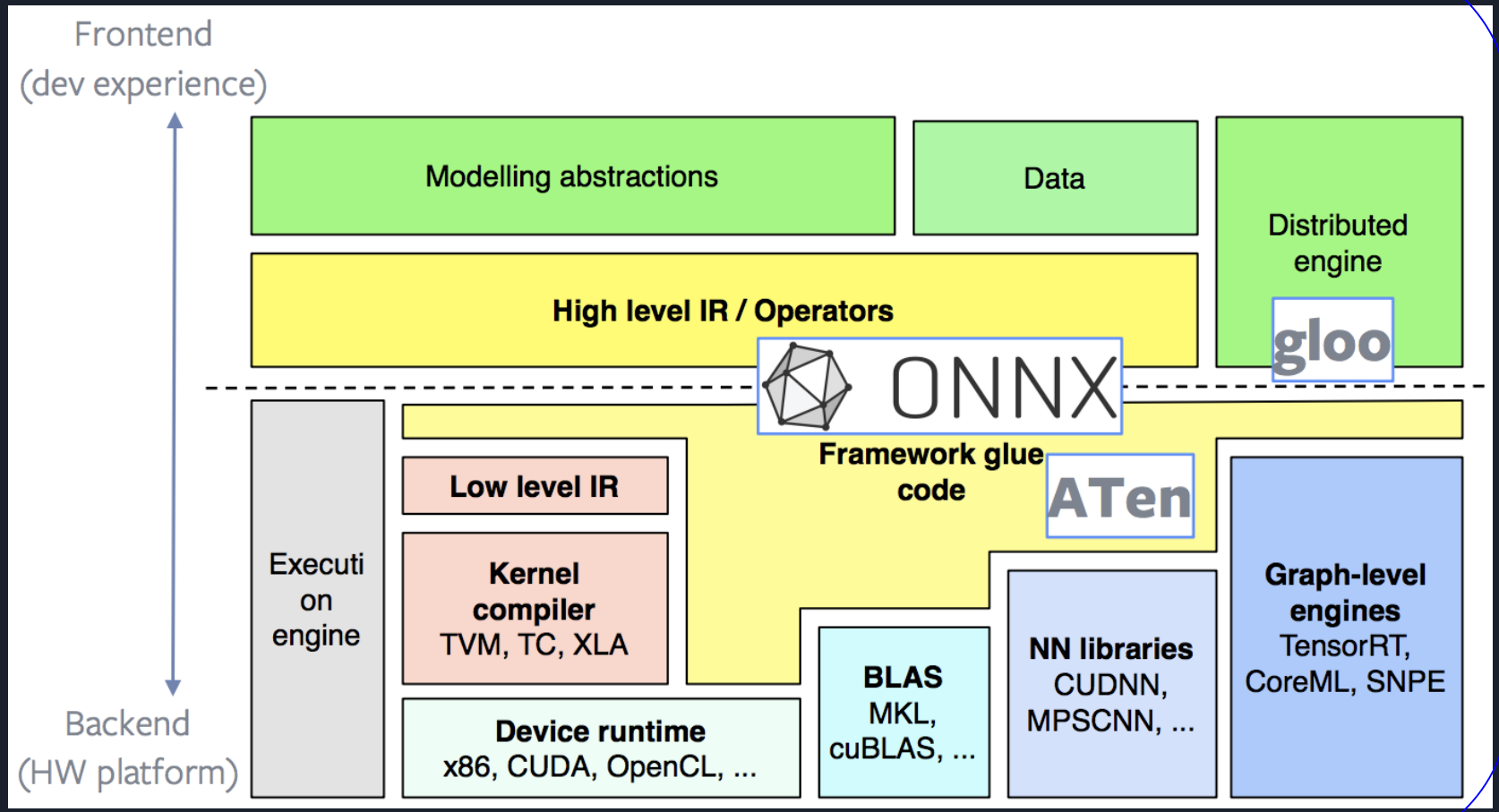


ONNX as the standard intermediate layer

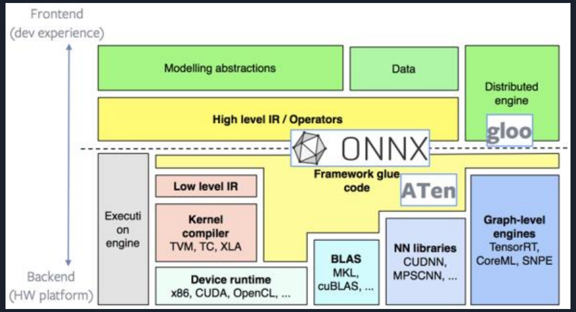


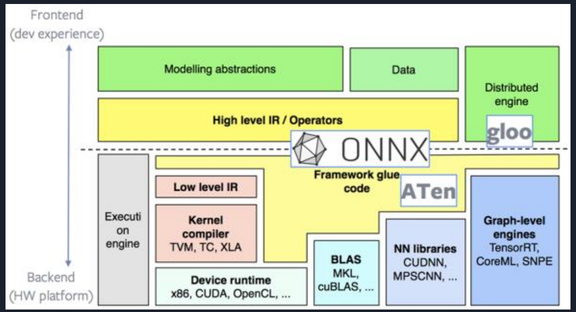
ONNX Help To Bridge, but in a cloud ...

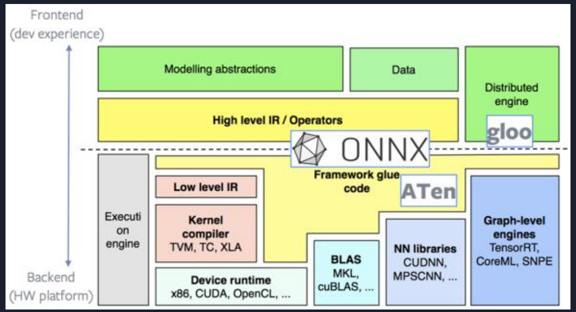
How can you effectively schedule each one of these onto the right node ?



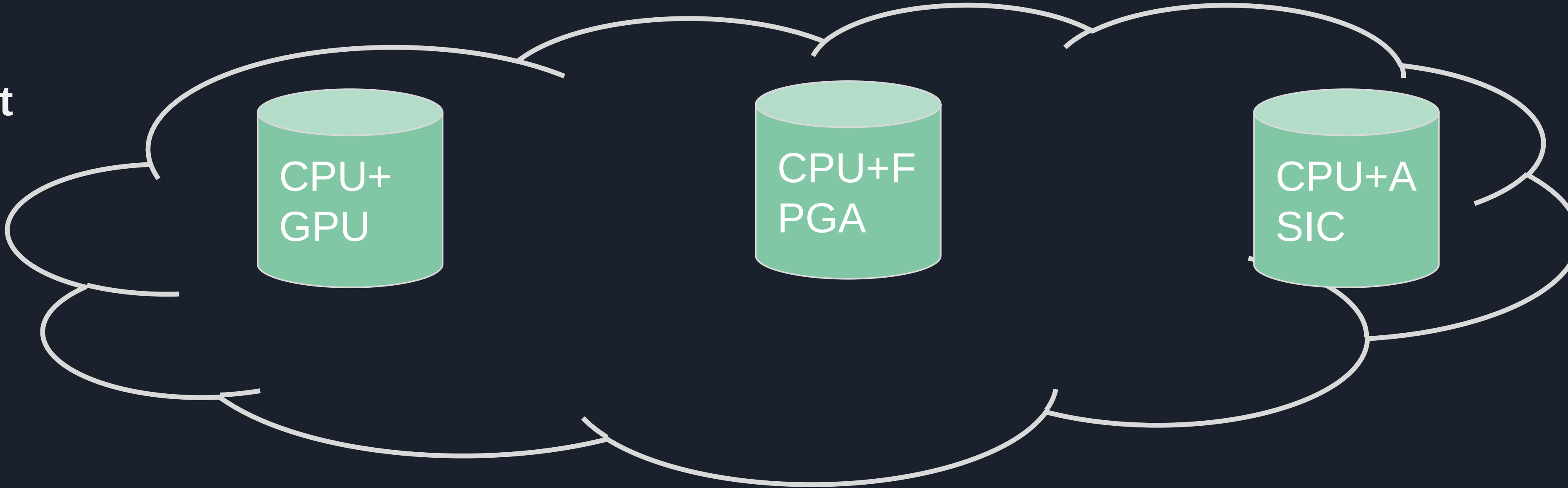
ONNX Help To Bridge, but in a cloud which you have to build ...







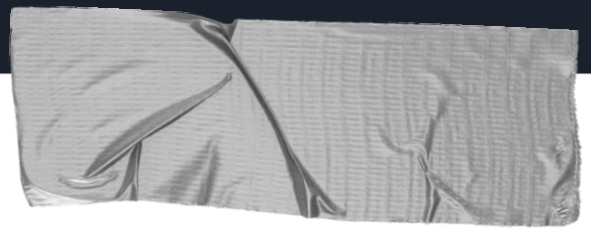
How can you effectively manage these nodes that run your framework ?



Asking The Right Questions

Can we have an AI cloud infrastructure software which

(1) Provides nice abstraction and management of the heterogeneous resources
(2) Is open source and driven by an open community
(3) Facilitates the e2e AI development



Outline

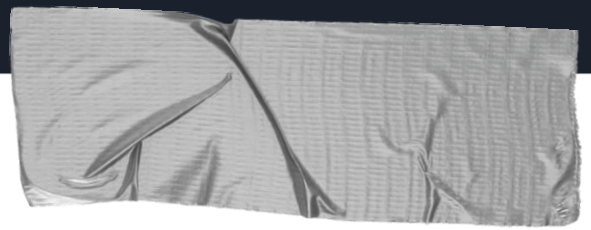
- ➔ **Challenges For AI**
- ➔ **Identify The Gap**
- ➔ **Introduce Cyborg Project**
- ➔ **Model Meets Resource**

Cyborg is a general management framework for accelerators

Proud OpenStack Official Project since 2017.09
(https://github.com/openstack/cyborg)

What can Cyborg offer now

(1) Discover Heterogeneous Resources(DSAs)
(2) Manage runtime libraries and packages
(3) Access and Quota control on resources
(4) Works with devices from various vendors
(5) Standardized APIs for resource management
(6) Program FPGA bitstreams through API portal
(7) Work with Nova Placement to schedule resources for VMs



Outline

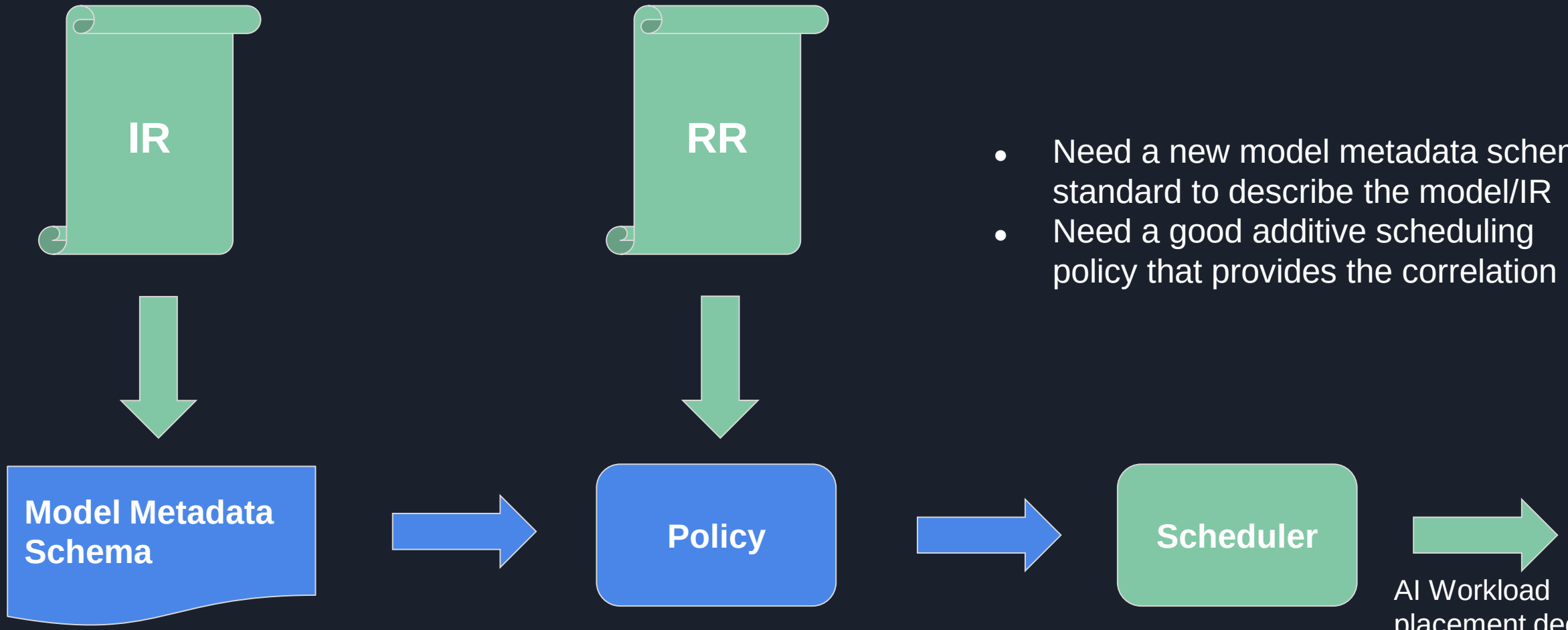
- ➔ **Challenges For AI**
- ➔ **Identify The Gap**
- ➔ **Introduce Cyborg Project**
- ➔ **Model Meets Resource**

ONNX + Cyborg

(1) ONNX Provides High Level Intermediate Representation(IR)
  (a) It answers the question of What the users want to do

(1) Cyborg Provides Resource Representation(RR)
  (a) It answers the question of How to do it

Model Meets Resource: Correlating IR and RR



- Need a new model metadata schema standard to describe the model/IR
- Need a good additive scheduling policy that provides the correlation

Model Meets Resource: Close The Gap

PaaS

Kubeflow + Kubernetes + Cyborg

IaaS

OpenStack Cyborg

Hardware
(CPU,GPU,FPGA,ASIC,...)

Heterogeneous Physical Resource

Agenda

01 Medical Background

02 Scenarios

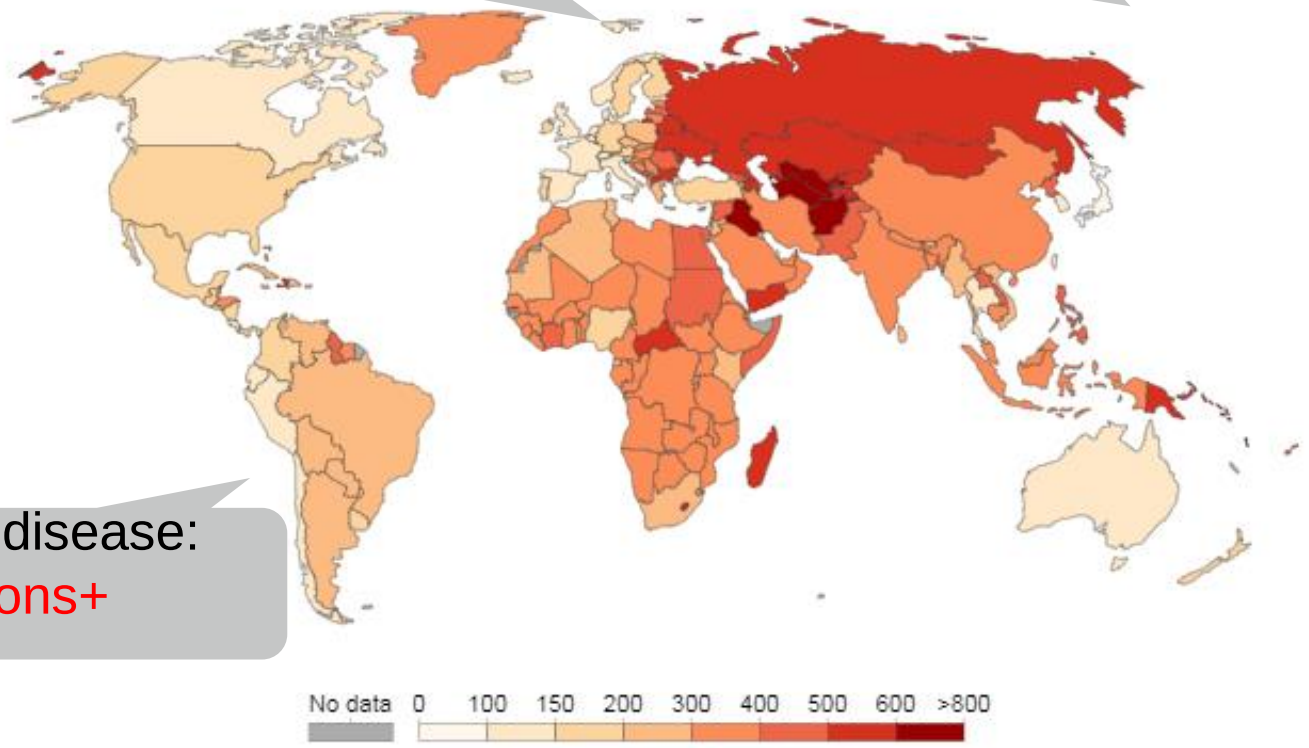
03 Edge Platform Introduction

1. Medical Background

Heart Diseases World Wide

CVD(cardiovascular disease)
30%+, 17 millions+

75%+ CVD deaths
In developing countries



Heart disease:
7 millions+

Medical resources are scarce and distributed unevenly.

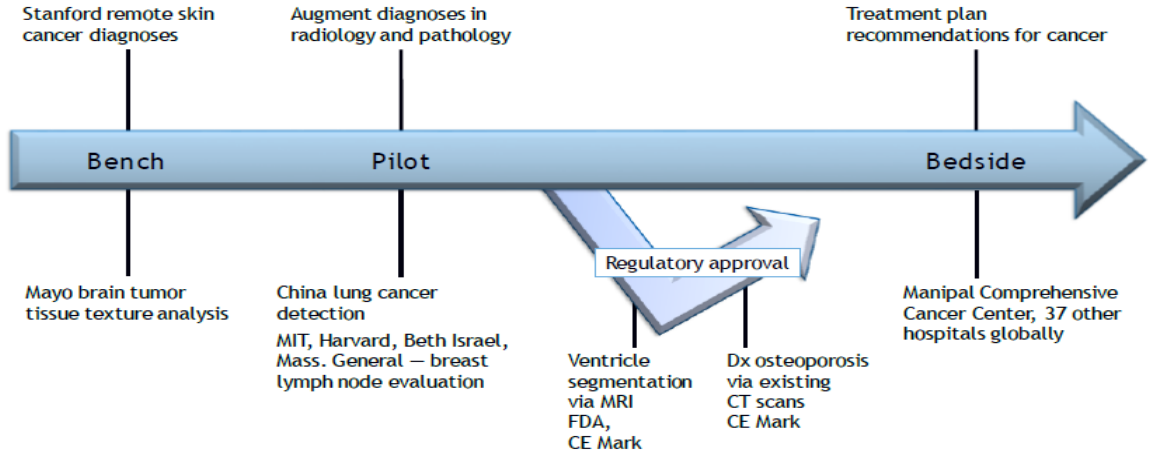
Deep Learning in Healthcare

- **Help in:**
  - Clinician augmentation
  - Prediction of risks of diseases

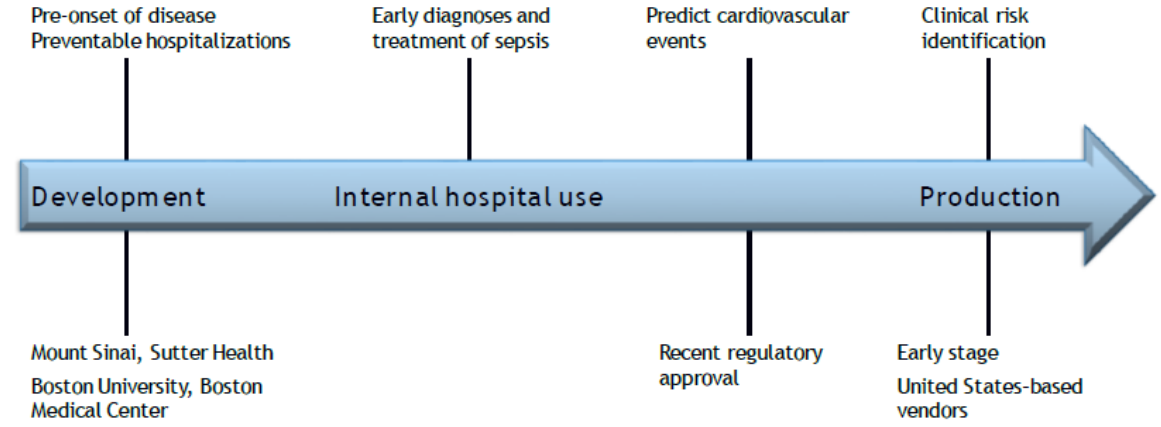
- **Advantages:**
  - Improve accuracy and speed to diagnoses and treatment
  - Address limited access to doctors

Train DL models with accurate data labeled by experienced physicians. Then help diagnosing using the trained DL models in clinics.

From Bench to Bedside – The Journey of Deep Learning to Assist Clinicians Diagnose and Treat

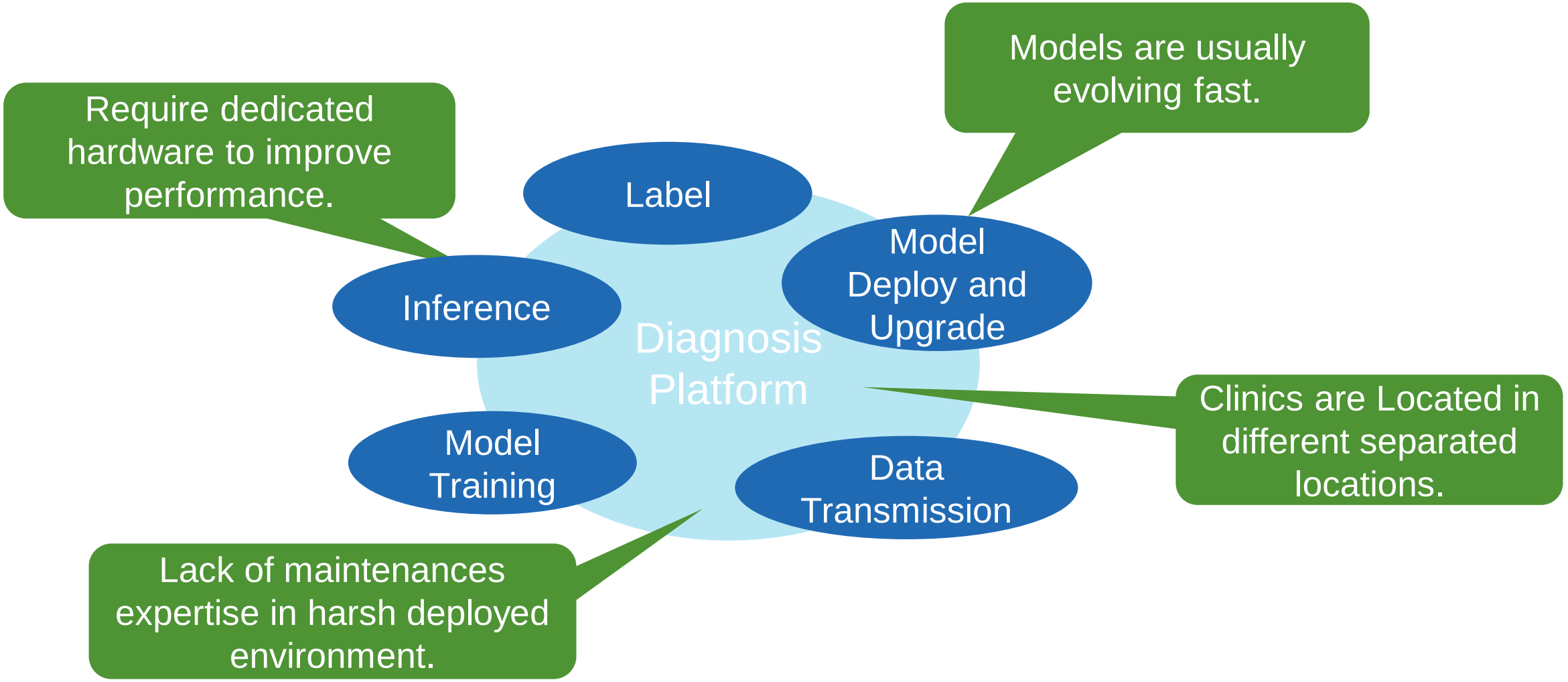


The Journey of Deep Learning in Prediction in Healthcare



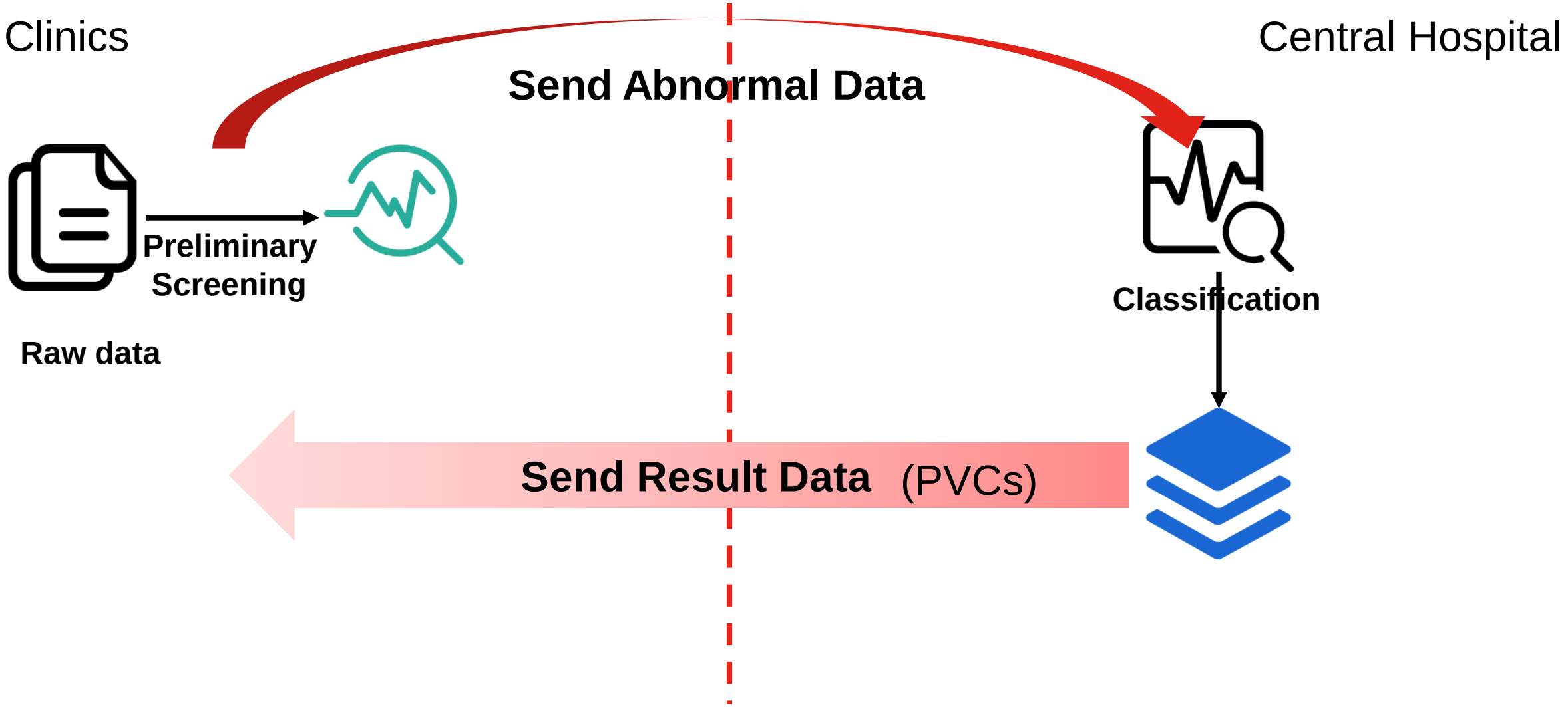
Source: IDC, 2017

Challenges

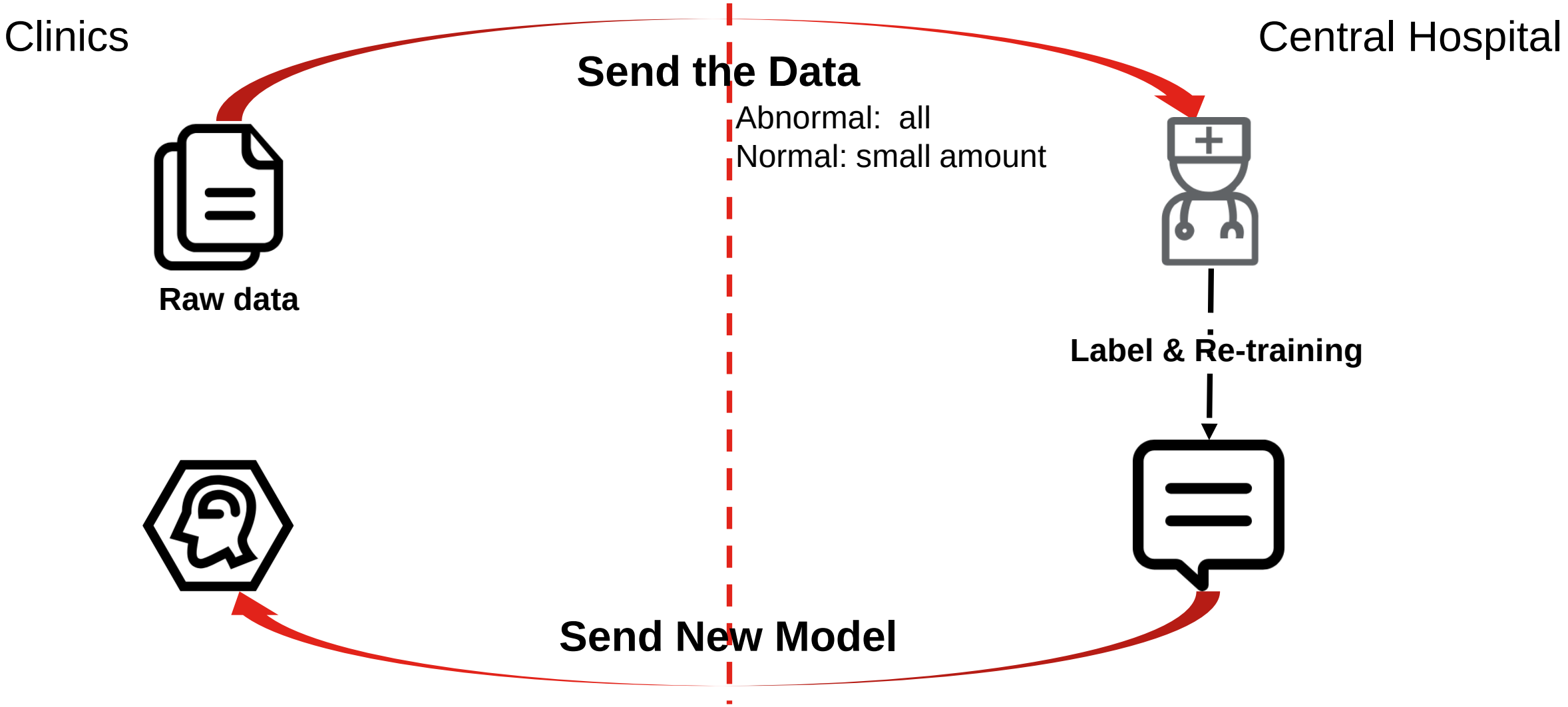


2. Scenarios

Diagnosis Process



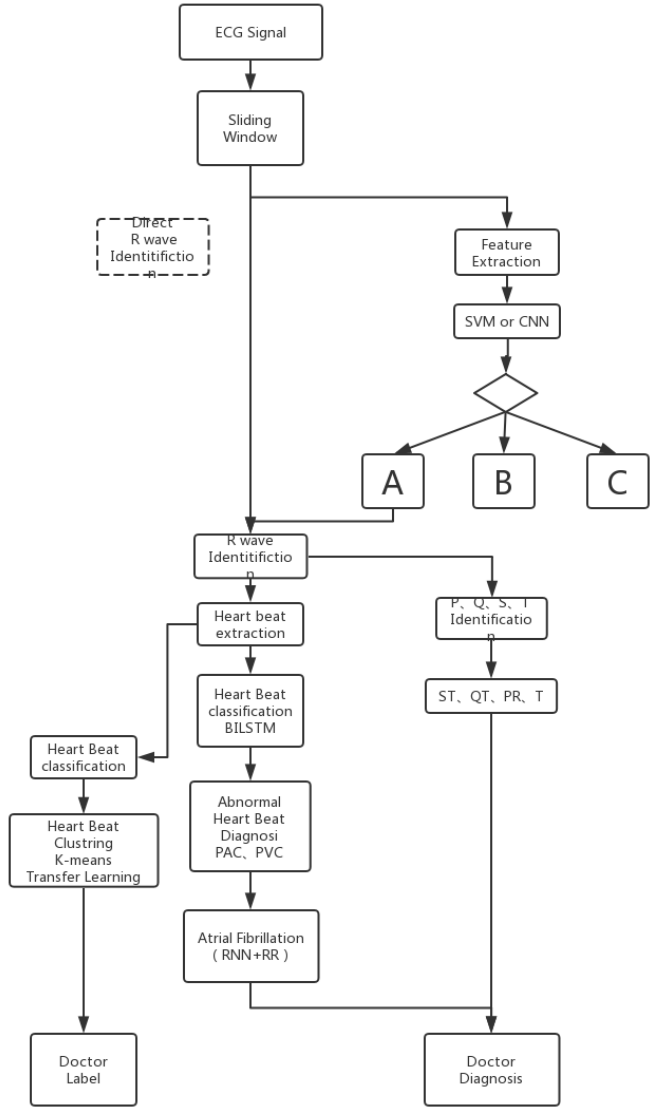
Dynamic Training Process



ECG Algorithms in the Edge & Cloud

- Edge Side: Preliminary Screening
  - Data normalization and de-noising
  - Signal quality assessment (SVM or CNN)
  - QRS waveform detection
  - Adaptive filtering
  - Abnormal/normal ECG signal detection

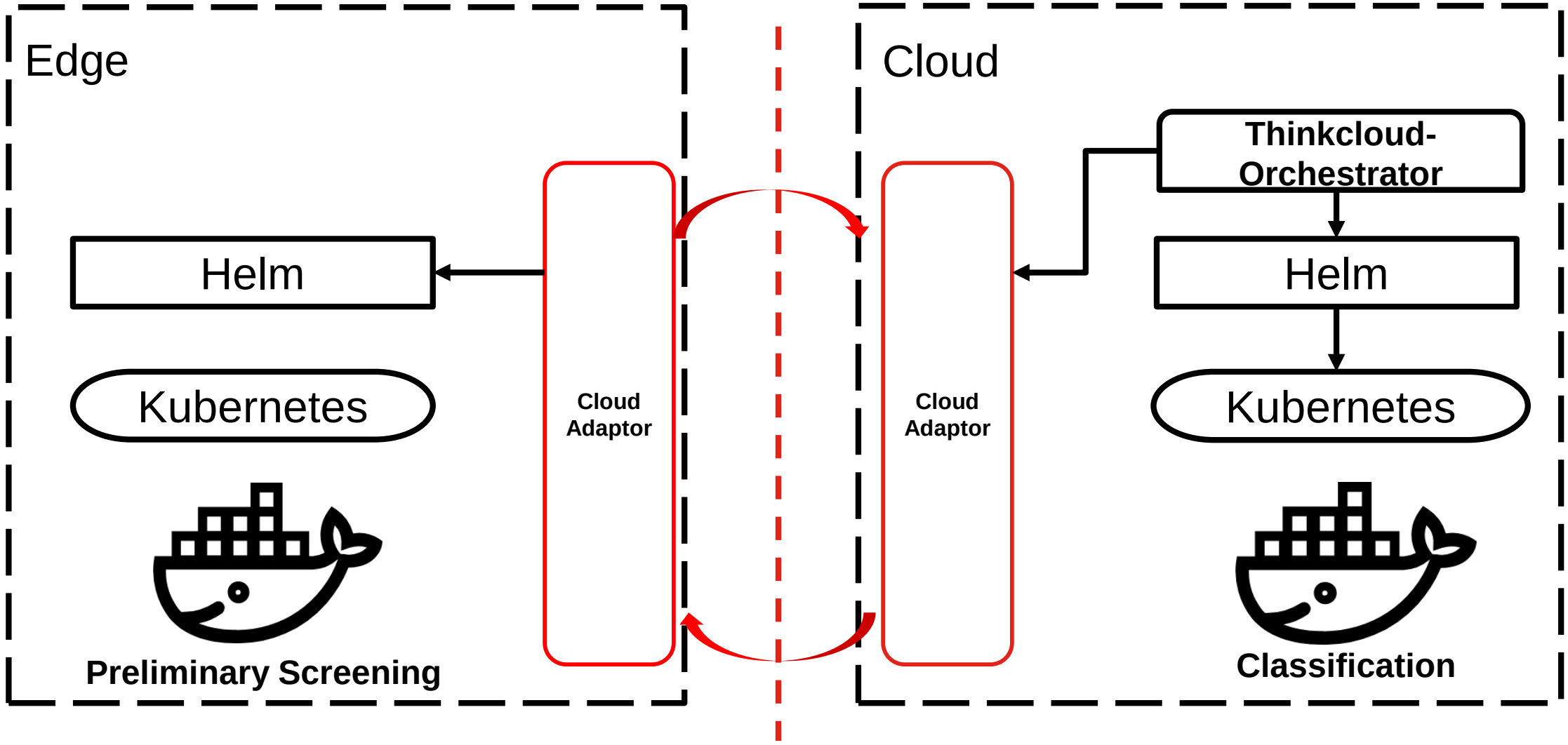
- Cloud Side: Classification
  - Entire processing flow of the ECG algorithm (K-means transfer learning, BilSTM, RNN+ RR)
  - Abnormal ECG signal classification
    - Premature ventricular contractions (PVCs)
    - Supraventricular premature beats
    - Atrial fibrillation, etc.



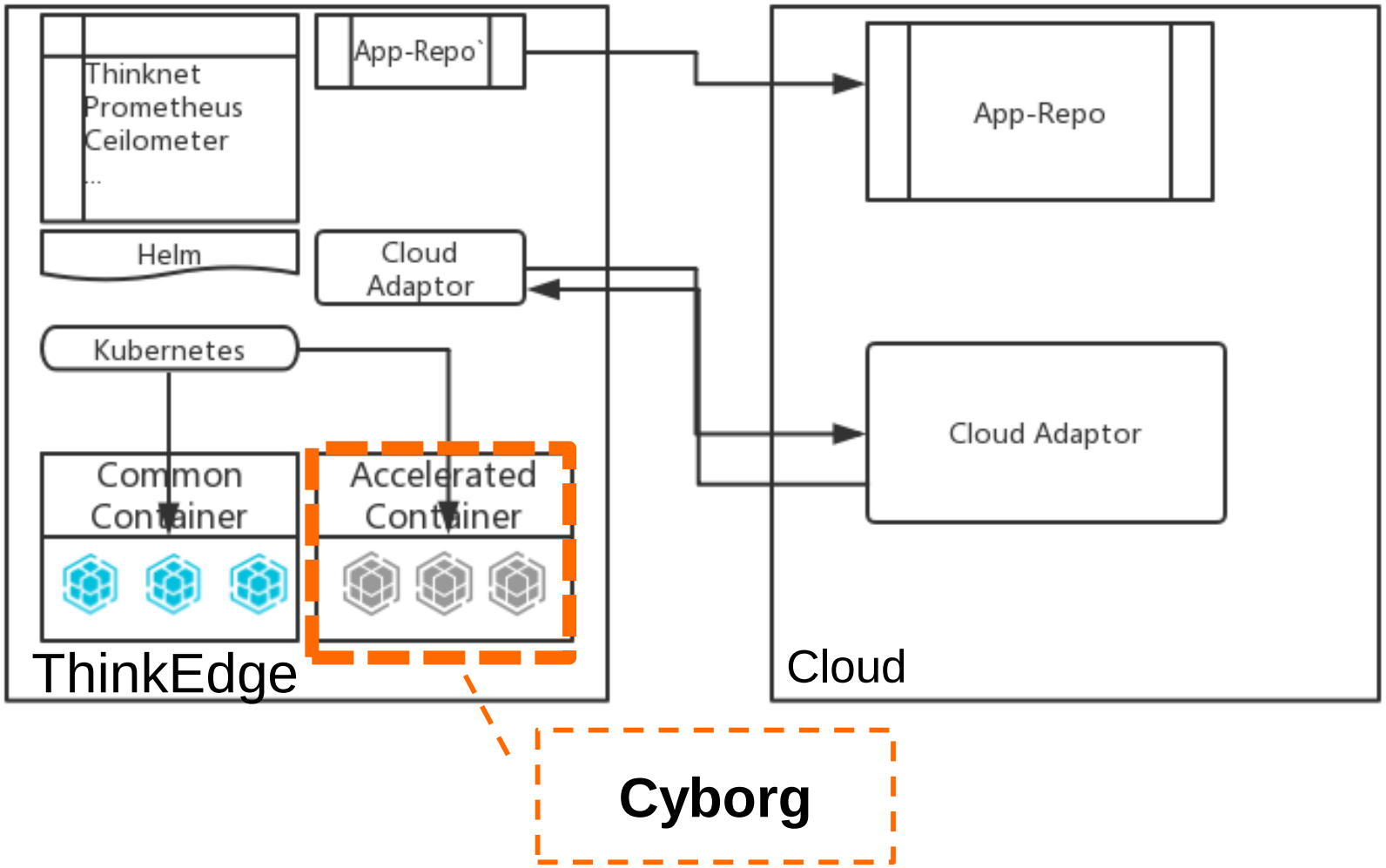
Classification Flow Diagram

3. Edge Platform Introduction

Cloud-Edge Collaboration

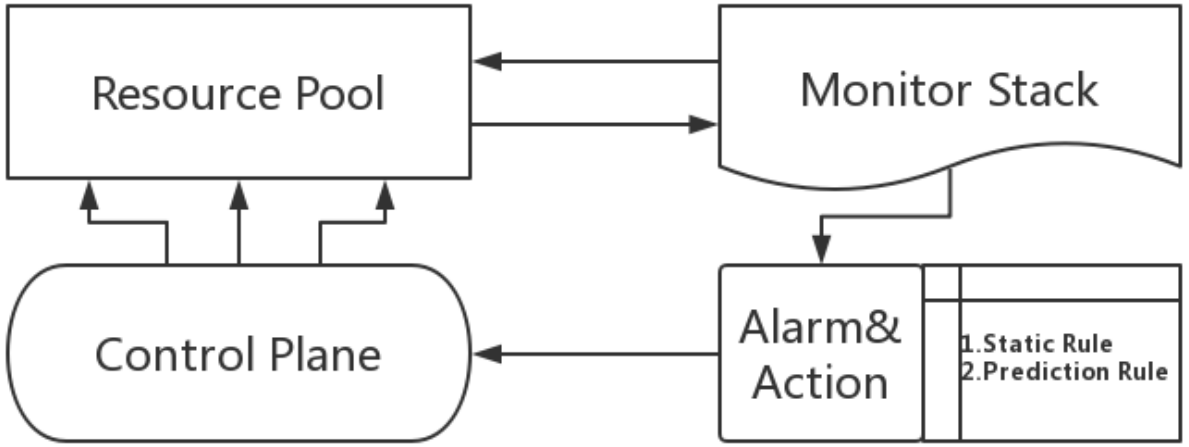


Application Acceleration

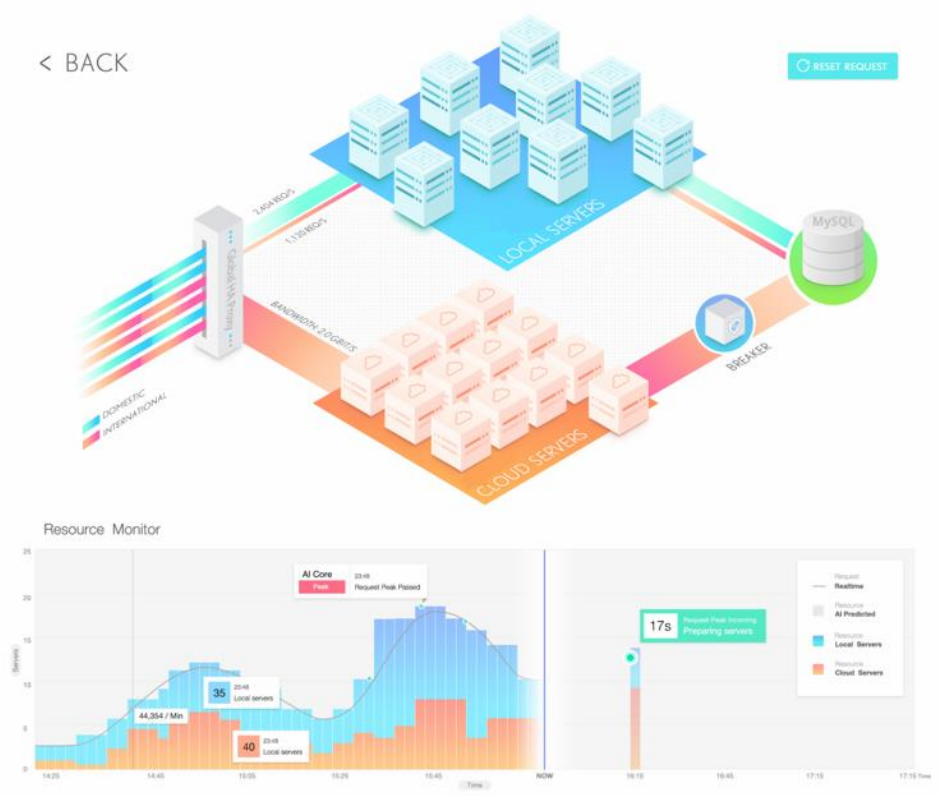


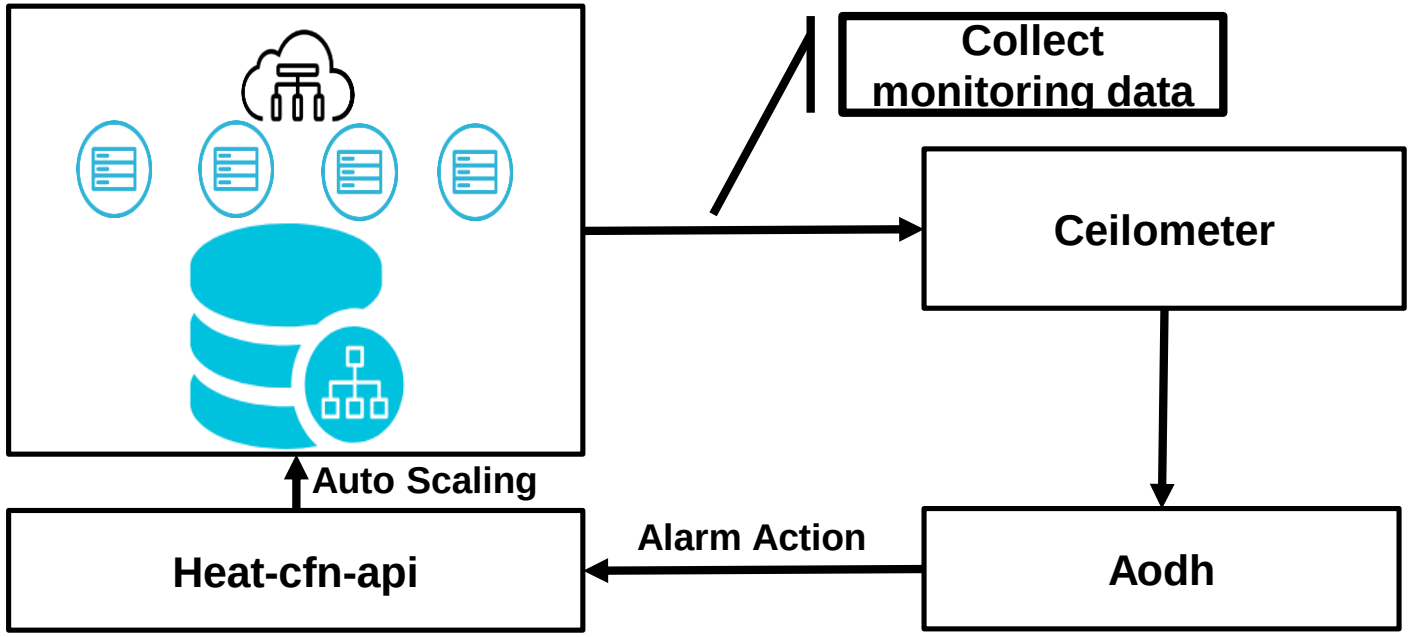
Workload Optimization & Monitoring

- Keyword
  - Automatic workload optimization
  - AI prediction workload optimization



Workload Optimization & Monitoring





Edge Platform Overview

Edge APP

Diseases Diagnosis APP

Edge Services

Message Queue

Edge-Cloud Data Sync

Load Balancer

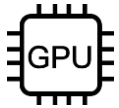
Infra Mgmt

Virtualization

Monitoring

Distributed Storage(Ceph)

Hardware

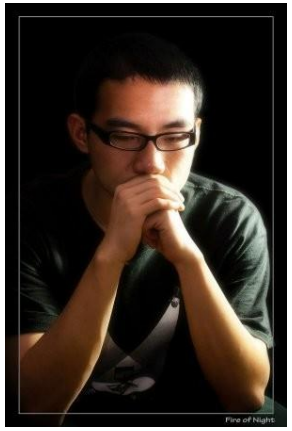


Mandatory

optional

Q&A

- Li Liu
  - Email: liliueecg@gmail.com
  - Senior Engineer in Huawei Canada Heterogeneous Cloud Acceleration Technology Lab
  - Project lead for OpenStack Cyborg(R and S releases) project
  - Toronto Raptor fan for years



- Jinghua Gao
  - Email: gaojh4@lenovo.com
  - Twitter: @Miss_Coco_Gao
  - OpenStack Cyborg Core Reviewer
  - Lenovo research, staff researcher

- Zhenghao Wang
  - Email: wangzh21@lenovo.com
  - IRC: wangzhh
  - OpenStack Cyborg Core Reviewer
  - Lenovo research, advisory researcher

thanks.

Lenovo 联想